ACADEMIC PRESS

Available at WWW.MATHEMATICSWEB.ORG POWERED BY SCIENCE DIRECT®

Journal of Multivariate Analysis 85 (2003) 130–155

Journal of Multivariate Analysis

http://www.elsevier.com/locate/jmva

On same-realization prediction in an infinite-order autoregressive process

Ching-Kang Ing[a,*] and Ching-Zong Wei[b]

[a] *Department of Statistics, National Taipei University, Taipei 104, Taiwan*
[b] *Institute of Statistical Science, Academia Sinica, Taipei 115, Taiwan*

Received 1 August 2001

Abstract

Let observations come from an infinite-order autoregressive (AR) process. For predicting the future of the observed time series (referred to as the *same-realization* prediction), we use the least-squares predictor obtained by fitting a finite-order AR model. We also allow the order to become infinite as the number of observations does in order to obtain a better approximation. Moment bounds for the inverse sample covariance matrix with an increasing dimension are established under various conditions. We then apply these results to obtain an asymptotic expression for the mean-squared prediction error of the least-squares predictor in same-realization and increasing-order settings. The second-order term of this expression is the sum of two terms which measure both the goodness of fit and model complexity. It forms the foundation for a companion paper by Ing and Wei (Order selection for same-realization predictions in autoregressive processes, Technical report C-00-09, Institute of Statistical Science, Academia Sinica, Taipei, Taiwan, ROC, 2000) which provides the first theoretical verification that AIC is asymptotically efficient for same-realization predictions. Finally, some comparisons between the least-squares predictor and the ridge regression predictor are also given.

AMS 1991 subject classifications: 62M10; 62M20

Keywords: Autoregressive process; Goodness of fit; Least squares; Model complexity; Ridge regression; Same-realization prediction

*Corresponding author.
E-mail addresses: cking@mail.ntpu.edu.tw (C.-K. Ing), czw@stat.sinica.edu.tw (C.-Z. Wei).

0047-259X/03/$ - see front matter
doi:10.1016/S0047-259X(02)00029-5

1. Introduction

Let observations $x_1, \dots, x_n$ come from a stationary autoregressive (AR) process $\{x_t\}$,

$$x_t + \sum_{i=1}^{\infty} a_i x_{t-i} = e_t, \quad t = \cdots, -1, 0, 1, \dots \tag{1.1}$$

where $\{e_t\}$ is a sequence of independent random variables with zero means and variances σ^2. In the literature, there are two different kinds of predictions under model (1.1). For *independent-realization* predictions, the aim is to predict another future independent series which has exactly the same probabilistic structure as the observed one. One of the special features of this type of prediction is that its mathematical analysis is relatively easy. This is because after conditioning on the observed series, the related prediction problem can be reduced to an estimation problem (see (4.3)). But for the practitioner, the emphasis is usually placed on *same-realization* predictions, that is, on the prediction of $x_{n+h},\ h \geqslant 1$. In the following, we concentrate on the performance of same-realization predictions. For a related discussion on independent-realization predictions, see [2–4,17].

When model (1.1) is a Gaussian AR process with finite known order, Fuller and Hasza [9] provided an expression for the mean-squared prediction error (MSPE) of the least-squares predictor of $x_{n+h},\ h \geqslant 1$, up to terms of order n^{-1}. For predicting x_{n+h} in general AR(∞) processes, Kunitomo and Yamamoto [15] used the least-squares predictor obtained by fitting a finite-order autoregressive (AR) model. They also gave an expression for the corresponding MSPE, up to terms of order n^{-1} in non-Gaussian settings.

Notice that when the underlying process is truly an AR(∞) process, all finite-order AR models are wrong. To reduce approximation errors, it is reasonable to increase the complexity (or the order) of the assumed model as more and more observations become available. This approach was first taken by Gerencsér [10] in the same-realization setting. He used the ridge regression predictor with increasing (AR) order to forecast x_{n+1}, and obtained an expression for the corresponding MSPE (see (4.5)). The second-order term in his expression can explain how the MSPE is affected by the adopted model. However, the conditions imposed on the AR coefficients and the order's increasing rate are rather restrictive. As a result, he cannot obtain a similar expression for the MSPE of the optimal ridge regression predictor in the situation where AR coefficients decay exponentially or algebraically (see the discussion preceding Corollary 1 in Section 4). In addition, the performance of the least-squares predictor in same-realization and increasing-order settings is still left unanswered.

Our article provides resolutions the above questions. To fix ideas, let us first introduce some notations and assumptions. In the following, we assume that in model (1.1) the coefficients a_i's are absolutely summable, and the polynomial

$$A(z) = 1 + \sum_{i=1}^{\infty} a_i z^i \tag{1.2}$$

is bounded away from zero for $|z| \leqslant 1$. By Weiner's theorem ([19, p. 245]), these assumptions are equivalent to requiring that x_t has a one-sided infinite moving-average representation

$$x_t = \sum_{i=0}^{\infty} b_i e_{n-i}, \tag{1.3}$$

where b_i's are absolutely summable with $b_0 = 1$, and the polynomial

$$B(z) = A^{-1}(z) = 1 + b_1 z + b_2 z^2 + \cdots$$

is bounded away from zero for $|z| \leqslant 1$. The spectral density function of $\{x_t\}$, $f(\lambda)$, can be expressed as

$$f(\lambda) = \frac{\sigma^2}{2\pi} |A(e^{-i\lambda})|^{-2} = \frac{\sigma^2}{2\pi} |B(e^{-i\lambda})|^2. \tag{1.4}$$

For predicting x_{n+1}, we consider finite-order approximations. For each stage n, let models $\mathrm{AR}(1), \ldots, \mathrm{AR}(K_n)$ be the family of approximation models, where K_n increases to infinity with n at a rate to be specified later. When a model $\mathrm{AR}(k), 1 \leqslant k \leqslant K_n$, is considered, we use $\hat{\mathbf{a}}(k)$ to estimate the model's coefficient vector, where

$$-\hat{\mathbf{a}}(k) = (\hat{a}_1(k), \ldots, \hat{a}_k(k))' = \hat{R}_n^{-1}(k) \frac{1}{N} \sum_{j=K_n}^{n-1} \mathbf{x}_j(k) x_{j+1}, \tag{1.5}$$

provided that the inverse of

$$\hat{R}_n(k) = \frac{1}{N} \sum_{j=K_n}^{n-1} \mathbf{x}_j(k) \mathbf{x}_j'(k) \tag{1.6}$$

exists. Here, $\mathbf{x}_j(k) = (x_j, \ldots, x_{j-k+1})'$ and $N = n - K_n$. We note that the difference between $\hat{\mathbf{a}}(k)$ and the least-squares estimator $\hat{\mathbf{a}}^L(k)$ will be asymptotically negligible under the assumptions on K_n and x_t we are using, where

$$-\hat{\mathbf{a}}^L(k) = \left(\sum_{j=k}^{n-1} \mathbf{x}_j(k) \mathbf{x}_j'(k) \right)^{-1} \sum_{j=k}^{n-1} \mathbf{x}_j(k) x_{j+1}.$$

Throughout this paper, we consider only $\hat{\mathbf{a}}(k)$ in order to simplify the discussion. The resulting one-step-ahead predictor is $\hat{x}_{n+1}(k) = -\mathbf{x}_n'(k)\hat{\mathbf{a}}(k)$, which will be referred to as the least-squares predictor. Our goal is to find an asymptotic expression for the MSPE of $\hat{x}_{n+1}(k)$, namely, $E[\{x_{n+1} - \hat{x}_{n+1}(k)\}^2]$ with $1 \leqslant k \leqslant K_n$.

As observed in (1.5) and (1.6), the major difficulty for this task lies in the fact that as $k \to \infty$, the dimension of $\hat{R}_n^{-1}(k)$ becomes infinite. For related finite-dimensional results (i.e., $K_n = K$ being fixed with n), Fuller and Hasza [9], under a Gaussian assumption on $\{e_t\}$, showed that for any $q > 0$,

$$E||\hat{R}_n^{-1}(K)||^q = O(1), \tag{1.7}$$

where $||C||^2 = \lambda_{\max}(C'C)$ is the maximum eigenvalue of the matrix $C'C$. Bhansali and Papangelou [5] and Papangelou [16] extended this identity to more general cases. However, their bounds for the left-hand side of (1.7) involve some unbounded functions of K. Their bounds no longer guarantee (1.7), when K is replaced by K_n and K_n tends to infinity with n. See Remark 2 in Section 2 for further details.

To tackle this difficulty, in Section 2, we establish some sharper upper bounds for $E||\hat{R}_n^{-1}(K_n)||^q$ and $E||\hat{R}_n^{-1}(K_n) - R^{-1}(K_n)||^q$, where $q > 0$, and $R(k) = E(\mathbf{x}_n(k)\mathbf{x}_n'(k))$. In particular, we show in Theorem 2 that $E||\hat{R}_n^{-1}(K_n)||^q = O(1)$ even if K_n increases to infinity with n. Armed with the results established in Section 2, an asymptotic expression for the MSPE of $\hat{x}_{n+1}(k)$, which holds uniformly for all $1 \leqslant k \leqslant K_n$, is obtained in Theorem 3 of Section 3. The second-order term of this expression is the sum of two terms. The first term, $(k/N)\sigma^2$, proportional to the order of the assumed model, can be viewed as a measure of model complexity. The second term, $||\mathbf{a} - \mathbf{a}(k)||_R^2$ (see (3.3)), which decreases as k increases, measures the goodness of fit for the assumed model.

In Section 4, we compare the MSPEs of least-squares predictors in same- and independent-realization settings. We show in Theorem 4 that these MSPEs have the same asymptotic expression. One important implication of this result is that it offers a theoretical basis for believing that a model with a small MSPE in an independent-realization setting will also perform well in a same-realization setting. For independent-realization predictions, Shibata [17] showed that the MSPE of the least-squares predictor with order selected by AIC [1] will ultimately achieve the best compromise between $(k/N)\sigma^2$ and $||\mathbf{a} - \mathbf{a}(k)||_R^2$. Motivated by Shibata's result and the asymptotic equivalence just mentioned, we show in a companion paper [13] that AIC still possesses a similar property for same-realization predictions. To the authors' knowledge, this is the first result that confirms AIC's validity in same-realization settings. It is also worth noting that the asymptotic equivalence between these two types of MSPEs should not be taken for granted because some recent studies showed that this equivalence does not hold in some other situations. (For more details, see the discussion after Theorem 4.)

In addition, we also apply the analysis for the least-squares predictor to improve the result of the ridge regression predictor given by Gerencsér [10]. Under rather mild conditions, we show in Corollary 1 that the MSPEs of the ridge regression and least-squares predictors have the same asymptotic expression. In the end of Section 4, a multi-step-ahead generalization of the result obtained in Theorem 3 (which focuses on a one-step-ahead prediction) is also given.

2. Moment bounds

Before establishing moment bounds for $||\hat{R}_n^{-1}(K_n)||^q, \ q > 0$, we first introduce the first moment bound theorem given by Findley and Wei [7]. The univariate version of this theorem is frequently used in the present article. Let $\{u_t\}$ and $\{v_t\}$ be stationary real-valued time series with autocovariance functions, $\gamma_u(\cdot)$ and $\gamma_v(\cdot)$, respectively. We shall assume that $\{u_t\}$ and $\{v_t\}$ have linear representations,

$$u_t = \sum_{j=-\infty}^{\infty} c_{1,j}\alpha_{t-j}, \tag{2.1}$$

and

$$v_t = \sum_{j=-\infty}^{\infty} c_{2,j}\beta_{t-j}, \tag{2.2}$$

where α_t and β_t are $\mathscr{F}_t$-measurable, and $\mathscr{F}_t, -\infty < t < \infty$ is an increasing sequence of σ-fields of events. Further, we also assume that α_t and β_t have the following properties, all with probability 1:

(M1) $E(\alpha_t|\mathscr{F}_{t-1}) = 0;\ E(\beta_t|\mathscr{F}_{t-1}) = 0.$
(M2) $E(\alpha_t\beta_t|\mathscr{F}_{t-1}) = \sigma_{\alpha\beta},\quad E(\alpha_t^2|\mathscr{F}_{t-1}) = \sigma_\alpha^2,\ E(\beta_t^2|\mathscr{F}_{t-1}) = \sigma_\beta^2.$
(M3) There is a finite constant C_p such that (2.3) holds for some $p \geqslant 1$,

$$\sup_{-\infty<t<\infty} E(\alpha_t^{4p}|\mathscr{F}_{t-1}) \leqslant C_p,$$

$$\sup_{-\infty<t<\infty} E(\beta_t^{4p}|\mathscr{F}_{t-1}) \leqslant C_p. \tag{2.3}$$

Consider a constant-coefficient quadratic form, $Q \equiv \sum_{s,t=1}^{T} u_s q(s,t) v_t$, where $q(s,t)$ are constants. The following theorem establishes a bound for the 2pth moment norm of Q.

Theorem 1 (First moment bound theorem). *Suppose the process α_t and β_t satisfy* (M1)–(M3), *with $p \geqslant 1$ in* (2.3). *Then there exists a constant K_p, depending only on p, $\sigma_{\alpha\beta}$, σ_α^2, σ_β^2, and C_p such that*

$$||Q - EQ||_{2p} \leqslant K_p\left(\sum_{s,t,l,w=1}^{T} q(s,t)q(l,w)\gamma_u(s-l)\gamma_v(t-w)\right)^{1/2}, \tag{2.4}$$

where $\gamma_u(j) = Eu_tu_{t-j}$, $\gamma_v(j) = Ev_tv_{t-j}$, $j = 0, \pm 1, \ldots,$ and for random variable z, $E||z||_{2p} = \{E|z|^{2p}\}^{1/2p}$.

The following assumptions on the underlying process, $\{x_t\}$, and the corresponding noise process, $\{e_t\}$, are also essential to our analysis.

(K.1) $\{x_t\}$ is a stationary process satisfying (1.1) and (1.2).
(K.2) Let the distribution function of e_t be denoted by F_t. There exist some positive

real numbers α, δ, and M such that

$$|F_t(x) - F_t(y)| \leqslant M|x-y|^{\alpha}$$

holds for all t, provided $|x-y| \leqslant \delta$.

(K.3) $\sup_{-\infty<t<\infty} E|e_t|^s < \infty, \ s = 1, 2, \ldots.$

Remark 1. (1) Condition (K.1) implies that the spectral density function $f(\lambda)$ of x_t satisfies $f_1 \leqslant f(\lambda) \leqslant f_2$ for some $0 < f_1 \leqslant f_2 < \infty$, where $-\pi \leqslant \lambda \leqslant \pi$. This fact also ensures that $\sup_{k \geqslant 1} ||R(k)|| < \infty$ and $\sup_{k \geqslant 1} ||R^{-1}(k)|| < \infty$.

(2) Assumption (K.2) is a rather mild condition from a practical point of view. For example, in the situation where e_t's are identically distributed, (K.2) is easily fulfilled by any distribution function with bounded density. It was first proposed by Papangelou [16] to replace condition (III) in [5]. This latter condition is difficult to verify except for Gaussian processes. For a multivariate generalization of (K.2), see [8, Section 4]).

Lemma 1 below establishes an upper bound for $E\lambda_{\min}^{-q}(\hat{R}_n(K_n))$ with $q > 0$, where $\lambda_{\min}(A)$ denotes the minimum eigenvalue of matrix A. In the sequel, we use C to denote a generic positive constant independent of sample size n and of any index with an upper (or lower) limit depending on n. But C may depend on the distributional properties of x_t. It also may have different values in different places.

Lemma 1. *Let $\{K_n\}$ be a sequence of positive integers satisfying $K_n = o(n^{1/2})$. Assume* (K.1) *and* (K.2). *Then, for any $q > 0$,*

$$E\lambda_{\min}^{-q}(\hat{R}_n(K_n)) = O(K_n^{(2+\theta)q}) \tag{2.5}$$

for all $\theta > 0$.

Proof. Define $\mathbf{x}_j = \mathbf{x}_j(K_n) = (x_j, \ldots, x_{j-K_n+1})'$, and

$$A = \begin{pmatrix} 1 & a_1 & \cdots & & a_{K_n-1} \\ 0 & 1 & & & \vdots \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & a_1 \\ 0 & \cdots & \cdots & 0 & 1 \end{pmatrix}.$$

Consider the following transformation of $\mathbf{x}_j$, $\boldsymbol{\phi}_j = A\mathbf{x}_j = \mathbf{e}_j + \boldsymbol{\eta}_j$, where $\mathbf{e}_j = (e_j, \ldots, e_{j-K_n+1})'$, and $\boldsymbol{\eta}_j = (\eta_{j_1}, \ldots, \eta_{j_{K_n}})'$ with $\eta_{j_i}, 1 \leqslant i \leqslant K_n$, being a linear combination of e_l, $l \leqslant j - K_n$.

It is not difficult to check the following facts:

(F1) $\boldsymbol{e}_j$ is independent of $\{\boldsymbol{\eta}_{l_1}, \boldsymbol{e}_{l_2}$ for $l_1 \leqslant j$, and $l_2 \leqslant j - K_n\}$,

(F2) $\lambda_{\min}^{-1}(\sum_{j=K_n}^{n-1} \boldsymbol{x}_j \boldsymbol{x}_j') \leqslant \lambda_{\max}(A'A)\lambda_{\min}^{-1}(\sum_{j=K_n}^{n-1} \boldsymbol{\phi}_j \boldsymbol{\phi}_j')$, and

(F3) $\lambda_{\max}(A'A) = O(1)$,

where (F3) is ensured by $\sum_{i=1}^{\infty}|a_i| < \infty$. In view of (F2) and (F.3), (2.5) follows from

$$E\left(N^q \lambda_{\min}^{-q}\left(\sum_{j=K_n}^{n-1} \boldsymbol{\phi}_j \boldsymbol{\phi}_j'\right)\right) \leqslant C(K_n^{2+\theta})^q. \tag{2.6}$$

Let $q, \theta > 0$ be arbitrarily given. To obtain (2.6), it suffices to show that for sufficiently large n,

$$E\left\{\lambda_{\min}^{-q}\left(\sum_{i=l_0+1}^{tK_n+l_0} \boldsymbol{\phi}_{iK_n+j}\boldsymbol{\phi}_{iK_n+j}'\right)\right\} \leqslant C(K_n^{1+\theta})^q, \tag{2.7}$$

for all integers $0 \leqslant j \leqslant K_n - 1$ and $0 \leqslant l_0 \leqslant \lfloor N/K_n \rfloor - tK_n$, where t, independent of n, is some positive integer to be specified later, and for real number a, $\lfloor a \rfloor$ denotes the largest integer $\leqslant a$. The reason why (2.7) is sufficient for (2.6) to be true is explained as follows. By the convexity of x^{-q}, $x > 0$, and some elementary matrix algebra, one has

$$N^q \lambda_{\min}^{-q}\left(\sum_{j=K_n}^{n-1} \boldsymbol{\phi}_j \boldsymbol{\phi}_j'\right) \leqslant \frac{1}{K_n} \sum_{j=0}^{K_n-1} \left(\frac{N}{K_n}\right)^q \lambda_{\min}^{-q}\left(\sum_{i=1}^{\lfloor N/K_n \rfloor} \boldsymbol{\phi}_{iK_n+j}\boldsymbol{\phi}_{iK_n+j}'\right). \tag{2.8}$$

The same reasoning also ensures that for $0 \leqslant j \leqslant K_n - 1$,

$$\begin{aligned}
&\left(\frac{N}{K_n}\right)^q \lambda_{\min}^{-q}\left(\sum_{i=1}^{\lfloor N/K_n \rfloor} \boldsymbol{\phi}_{iK_n+j}\boldsymbol{\phi}_{iK_n+j}'\right) \\
&\quad \leqslant \frac{C}{C_N} \sum_{s=0}^{C_N-1} (tK_n)^q \lambda_{\min}^{-q}\left(\sum_{i=1}^{tK_n} \boldsymbol{\phi}_{(i+stK_n)K_n+j}\boldsymbol{\phi}_{(i+stK_n)K_n+j}'\right),
\end{aligned} \tag{2.9}$$

where $C_N = \lfloor \lfloor N/K_n \rfloor / (tK_n) \rfloor$. In view of (2.8) and (2.9), (2.6) follows from (2.7).

In the following, we only prove (2.7) for the case of $l_0 = 0$ and $j = 0$ because other cases can be obtained similarly. To simplify the notation, define $\boldsymbol{z}_i = \boldsymbol{\phi}_{iK_n}$. Then, the left-hand side of (2.7) (with $l_0 = 0$ and $j = 0$) is bounded by

$$\begin{aligned}
&(g_n)^q + \int_{(g_n)^q}^{\infty} P(D_n(u))\,du + \int_{(g_n)^q}^{\infty} P(T_n(u) \cap D_n^C(u))\,du \\
&\quad \equiv (g_n)^q + V_{1n} + V_{2n},
\end{aligned} \tag{2.10}$$

where $g_n = (36/\delta^2)K_n^{1+\theta}$,

$$T_n(u) = \left\{\inf_{||\boldsymbol{y}||=1} \sum_{i=1}^{tK_n} (\boldsymbol{z}_i'\boldsymbol{y})^2 < u^{-1/q}\right\},$$

$$D_n(u) = \left\{\sum_{i=1}^{tK_n} ||\boldsymbol{z}_i||^2 > \frac{u^{2l/q}}{K_n}\right\}$$

with $l > (3+q)/2$, and $D_n^C(u)$ denotes the complement of $D_n(u)$. Notice that by Chebyshev's inequality, $V_{1n} \leqslant C$ for all $n \geqslant 1$. In view of this fact and (2.10), (2.7) (with $l_0 = 0$ and $j = 0$) is guaranteed by showing that for all large n,

$$V_{2n} \leqslant C.$$

To deal with V_{2n}, first consider the hypersphere $S_n = \{\boldsymbol{y}\colon ||\boldsymbol{y}|| = 1\} \subset R^{K_n}$ and the hypercube $H^{K_n} = [1 - 2u^{-(l+1/2)q^{-1}}(\lfloor u^{(l+1/2)q^{-1}} \rfloor + 1), 1]^{K_n}$, where $u \geqslant (g_n)^q$. Since $u^{-(l+1/2)q^{-1}}(\lfloor u^{(l+1/2)q^{-1}} \rfloor + 1) \geqslant 1$, then $S_n \subseteq H^{K_n}$. Divide H^{K_n} into subhypercubes, each of which has an edge of length $2u^{-(l+1/2)q^{-1}}$ and a circumscribed hypersphere of radius $\sqrt{K_n} u^{-(l+1/2)q^{-1}}$. Let these subhypercubes be denoted by B_i for $1 \leqslant i \leqslant m^*$. Then, it can be seen that the number of B_i's, m^*, does not exceed $(\lfloor u^{(l+1/2)q^{-1}} \rfloor + 1)^{K_n}$. Write $S_n = \bigcup_{i=1}^{m^*}(S_n \cap B_i) \equiv \bigcup_{i=1}^{m^*} G_i$. Then,

$$P(T_n(u) \cap D_n^C(u)) \leqslant \sum_{j=1}^{m^*} P(Q_{nj}(u)), \tag{2.11}$$

where $Q_{nj}(u) = \bigcap_{i=1}^{tK_n}\{\inf_{\boldsymbol{y}\in G_j}|\boldsymbol{y}'\boldsymbol{z}_i| < u^{-1/(2q)}, ||\boldsymbol{z}_i|| \leqslant K_n^{-1/2}u^{l/q}\}$.

Since for any $\boldsymbol{l}_j, \boldsymbol{y}_j \in G_j,\ j = 1, \ldots, m^*$,

$$|\boldsymbol{l}_j'\boldsymbol{z}_i| \leqslant ||\boldsymbol{l}_j - \boldsymbol{y}_j||\,||\boldsymbol{z}_i|| + |\boldsymbol{y}_j'\boldsymbol{z}_i|,$$

then

$$|\boldsymbol{l}_j'\boldsymbol{z}_i| \leqslant 2u^{-(2q)^{-1}} + \inf_{\boldsymbol{y}\in G_j} |\boldsymbol{y}'\boldsymbol{z}_i| \leqslant 3u^{-(2q)^{-1}}$$

on the set

$$\left\{\inf_{\boldsymbol{y}\in G_j} |\boldsymbol{y}'\boldsymbol{z}_i| < u^{-(2q)^{-1}},\ ||\boldsymbol{z}_i|| \leqslant K_n^{-1/2}u^{lq^{-1}}\right\}.$$

This gives for any $\boldsymbol{l}_j \in G_j$ and all $1 \leqslant j \leqslant m^*$,

$$Q_{nj}(u) \subseteq \bigcap_{i=1}^{tK_n} \{|\boldsymbol{l}_j'\boldsymbol{z}_i| \leqslant 3u^{-1/(2q)}\}. \tag{2.12}$$

Let $E_{j,i} = \{|\boldsymbol{l}_j'\boldsymbol{z}_i| \leqslant 3u^{-(2q)^{-1}}\}$. Then by (F.1),

$$\begin{aligned} P\left(\bigcap_{i=1}^{tK_n} \{|\boldsymbol{l}_j'\boldsymbol{z}_i| \leqslant 3u^{-1/(2q)}\}\right) &= E\left(\prod_{i=1}^{tK_n} I_{E_{j,i}}\right) \\ &= E\left\{\prod_{i=1}^{tK_n-1} I_{E_{j,i}} P(|\boldsymbol{l}_j'\mathbf{e}_{tK_n^2} + \boldsymbol{l}_j'\boldsymbol{\eta}_{tK_n^2}| \leqslant 3u^{-(2q)^{-1}} | e_l, l \leqslant tK_n^2 - K_n)\right\}, \end{aligned} \tag{2.13}$$

where I_A denotes the indicator function for event A. Denote the ith component of $\boldsymbol{l}_j$ by $l_{j,i}$. Since $||\boldsymbol{l}_j||^2 = 1$, there is a positive integer $1 \leqslant s_j \leqslant K_n$ such that $|l_{j,s_j}| \geqslant (K_n)^{-1/2}$.

Without loss of generality, we also assume $l_{j,s_j} > 0$. Then, by (K.2) and (F.1), all with probability 1,

$$\begin{aligned}
&P(|\boldsymbol{l}_j'\mathbf{e}_{tK_n^2} + \boldsymbol{l}_j'\boldsymbol{\eta}_{tK_n^2}| \leqslant 3u^{-(2q)^{-1}} | e_l, l \leqslant tK_n^2 - K_n) \\
&\quad \leqslant P(-3u^{-(2q)^{-1}} + a \leqslant l_{j,s_j} e_{tK_n^2 - s_j + 1} \leqslant 3u^{-(2q)^{-1}} + a | e_l, l \leqslant tK_n^2 - K_n) \\
&\quad \leqslant M(K_n^{1/2} 6u^{-(2q)^{-1}})^{\alpha},
\end{aligned}$$

where

$$a = -\mathbf{l}_j'\boldsymbol{\eta}_{tK_n^2} - \sum_{\substack{i=1 \\ i \neq s_j}}^{K_n} l_{j,i} e_{tK_n^2 - i + 1}.$$

This fact and (2.13) yield

$$P\left(\bigcap_{i=1}^{tK_n} \{|\boldsymbol{l}_j'\boldsymbol{z}_i| \leqslant 3u^{-1/(2q)}\}\right) \leqslant E\left(\prod_{i=1}^{tK_n - 1} I_{E_{j,i}}\right) M(K_n^{1/2} 6u^{-(2q)^{-1}})^{\alpha}.$$

Repeating this argument $tK_n - 1$ times, the left-hand side of (2.13) is bounded by

$$(M6^{\alpha} K_n^{\alpha/2})^{tK_n} u^{-\alpha t K_n (2q)^{-1}}.$$

This bound, (2.11), (2.12) and the upper bound for m^* mentioned above imply that the left-hand side of (2.11) is dominated by

$$C(M6^{\alpha} K_n^{\alpha/2})^{tK_n} u^{-K_n(\alpha t - 2l - 1)(2q)^{-1}}. \tag{2.14}$$

Now, by taking $t \geqslant \max\{\lfloor (2l+1)(1+\theta)/(\alpha\theta) \rfloor, \lfloor (2l+1+2q)/\alpha \rfloor\} + 1$, $V_{2n} \leqslant C$ follows from integrating (2.14) over u, and hence the proof is complete. □

Remark 2. The proof for Lemma 1 was inspired by Bhansali and Papangelou [5]. It is also closely related to Papangelou [16] and Findley and Wei [8, Theorem 2], which deals with a multivariate version of (1.7) in non-Gaussian vector time series. However, a more delicate analysis is needed here to obtain a sharper bound. Note that if (K.2) holds and e_i's are i.i.d. random variables having finite moments of all orders, then following Papangelou [16, Theorem 3.2] and an argument of Bhansali and Papangelou [5, pp. 1159–1160], an upper bound for the left-hand side of (2.5) is given by

$$CK_n^{2q}(4K_n^2)^{K_n - 1} M^{\tilde{K}_n} \left(\sqrt{8K_n^2\left(1 + \sum_{j=1}^{\infty} a_j^2\right)(\tilde{K}_n K_n + 1)}\right)^{\alpha \tilde{K}_n}, \tag{2.15}$$

provided $K_n^2 < \gamma n$, for some positive number γ. Here, $\tilde{K}_n$ is a positive number satisfying $\alpha\tilde{K}_n > (4K_n + 2q - 2)$ with α defined in (K.2). When K_n is bounded by a finite positive number, (2.15) ensures that $E\lambda_{\min}^{-q}(\hat{R}_n(K_n))$ is also bounded. However, if K_n tends to infinity with n, then the boundedness of $E\lambda_{\min}^{-q}(\hat{R}_n(K_n))$ is no longer guaranteed by (2.15). Moreover, since for large K_n, (2.15) is not less than $(\eta K_n)^{10K_n}$ for some positive number η (independent of K_n), even if K_n increases to infinity at a very slow rate, (2.15) still provides an extremely large value. For example, if $K_n = \gamma_1 \log n,\ \gamma_1 > 0$, then it is easy to see that $n/(\eta K_n)^{10K_n} = o(1)$.

Equality (2.5) guarantees that $\hat{R}_n^{-1}(K_n)$ almost surely exists for all large n. Therefore, we are allowed to define $\hat{R}_n^{-1}(K_n)$ as any generalized inverse of $\hat{R}_n(K_n)$ without affecting the related asymptotic results. Hence, (2.5) can be rewritten as follows: for $q > 0$,

$$E||\hat{R}_n^{-1}(K_n)||^q = O(K_n^{(2+\theta)q}) \tag{2.16}$$

holds for all large n and any $\theta > 0$. Although the upper bound given by the right-hand side of (2.16) is still not bounded as K_n tends to infinity, its moderate value (in comparison with (2.15)) provides a foundation for further improvement. We now begin with the following lemma.

Lemma 2. *Assume* (K.1) *and* $\sup_{-\infty < t < \infty} E(|e_t|^{2\max\{q,2\}}) < \infty$, *for some* $q > 0$. *Then,*

$$E||\hat{R}_n(K_n) - R(K_n)||^q \leqslant C\left(\frac{K_n^2}{N}\right)^{q/2}. \tag{2.17}$$

Proof. We only prove (2.17) for $q \geqslant 2$ because this and Jensen's inequality can easily yield the result for $q < 2$. First observe that

$$E||\hat{R}_n(K_n) - R(K_n)||^q \leqslant \frac{K_n^q}{K_n^2} \sum_{i=1}^{K_n} \sum_{j=1}^{K_n} E|\hat{r}_{i,j} - r_{i-j}|^q, \tag{2.18}$$

where $\hat{r}_{i,j}$ and r_{i-j} denote the (i,j) components of $\hat{R}_n(K_n)$ and $R(K_n)$, respectively. We also have

$$E|\hat{r}_{i,j} - r_{i-j}|^q = N^{-q} E\left\{\left|\sum_{s=K_n}^{n-1} \sum_{v=K_n}^{n-1} q(s,v)(x_{s+1-i}x_{v+1-j} - r_{s-v+j-i})\right|^q\right\}, \tag{2.19}$$

where $q(s,v) = 1$ if $s = v$, otherwise $q(s,v) = 0$. By Theorem 1 and (K.1), the term on the right-hand side of (2.19) is bounded by

$$CN^{-q/2}\left(\sum_{i=-N+1}^{N-1} \left(1 - \left|\frac{i}{N}\right|\right) r_i^2\right)^{q/2} \leqslant CN^{-q/2}.$$

This bound and (2.18) yield (2.17). $\square$

The main result of this section given in the following theorem.

Theorem 2. (i) *Assume* (K.1), (K.2), $K_n^{6+\delta_1} = O(n)$ *for some* $\delta_1 > 0$, *and* $\sup_{-\infty < t < \infty} E(|e_t|^{2\max\{q_1,2\}}) < \infty$ *for some* $q_1 > 0$. *Then, for any* $0 < q < q_1$

$$E||\hat{R}_n^{-1}(K_n)||^q \leqslant C, \tag{2.20}$$

and

$$E||\hat{R}_n^{-1}(K_n) - R^{-1}(K_n)||^{q/2} \leqslant C\left(\frac{K_n^2}{N}\right)^{q/4}, \tag{2.21}$$

for sufficiently large n.

(ii) *Assume* (K.1)–(K.3), *and* $K_n^{2+\delta_1} = O(n)$ *for some* $\delta_1 > 0$. *Then, for any* $q > 0$, (2.20) *and* (2.21) *hold for sufficiently large* n.

Proof. To prove (i), first observe that by Lemma 1 and the definition of $\hat{R}_n^{-1}(K_n)$,

$$||\hat{R}_n^{-1}(K_n) - R^{-1}(K_n)||^q \leqslant ||\hat{R}_n^{-1}(K_n)||^q ||\hat{R}_n(K_n) - R(K_n)||^q ||R^{-1}(K_n)||^q \tag{2.22}$$

almost surely for large n. This fact, Hölder's inequality, (2.16), and Remark 1 ensure that for any $\theta > 0$,

$$E||\hat{R}_n^{-1}(K_n) - R^{-1}(K_n)||^q \leqslant C(E||\hat{R}_n(K_n) - R(K_n)||^{q_1})^{q/q_1}(K_n^{2+\theta})^q \tag{2.23}$$

for sufficiently large n.

According to (2.17) and (2.23), we have for large n,

$$E||\hat{R}_n^{-1}(K_n) - R^{-1}(K_n)||^q \leqslant C\left(\frac{K_n^{6+2\theta}}{N}\right)^{q/2}. \tag{2.24}$$

Set $2\theta \leqslant \delta_1$. Then, (2.20) follows from (2.24). Moreover, since the Cauchy–Schwarz inequality gives

$$E||\hat{R}_n^{-1}(K_n) - R^{-1}(K_n)||^{q/2} \leqslant C(E||\hat{R}_n^{-1}(K_n)||^q)^{1/2}(E||\hat{R}_n(K_n) - R(K_n)||^q)^{1/2}, \tag{2.25}$$

(2.21) follows from (2.20) and (2.17).

To prove (ii), we first arbitrarily choose $\theta > 0$. By (2.24) and Remark 1, one has for any $q > 0$,

$$E||\hat{R}_n^{-1}(K_n)||^q \leqslant C\left(1 + \left(\frac{K_n^{6+2\theta}}{N}\right)^{q/2}\right),$$

for all large n. This, (2.17), (2.25), and Remark 1 imply that

$$E||\hat{R}_n^{-1}(K_n)||^{q/2} \leqslant C\left(1 + \left(\frac{K_n^{8+2\theta}}{N^2}\right)^{q/4}\right).$$

Repeating this argument $s-1$ times, one has for all large n,

$$E||\hat{R}_n^{-1}(K_n)||^{q2^{-s}} \leqslant C\left(1 + \left(\frac{K_n^{2+\delta_2}}{N}\right)^{(1+s)q2^{-(s+1)}}\right), \tag{2.26}$$

where $\delta_2 = (4+2\theta)/(s+1)$. Set $s = \lfloor (4+2\theta)/\delta_1 \rfloor$. Then, $\delta_2 < \delta_1$. This and the hypothesis that $K_n^{2+\delta_1} = O(n)$ yield $K_n^{2+\delta_2}/N \to 0$, and hence the left-hand side of (2.26) is bounded by a finite positive number for all large n. Since q in (2.26) is arbitrary, (2.20) follows. Moreover, (2.21) is given by (2.20), (2.25), and (2.17). $\square$

Remark 3. If, in place of the right-hand side of (2.16), (2.15) is used as a bound, then with an argument similar to that used for verifying Theorem 2, (2.20) and (2.21) hold with a very stringent limitation on K_n, namely, $K_n \log K_n = O(\log n)$ (which implies that $K_n = o(\log n)$). Since K_n represents the order of the largest model in the approximating family, this limitation is problematic from a model selection point of view because it will ultimately rule out optimal models in many important situations (e.g., in which AR coefficients decay exponentially or algebraically). See Remark 5 for more details.

Inequality (2.20) immediately implies

$$\sup_{1 \leqslant k \leqslant K_n} E\|\hat{R}_n^{-1}(k)\|^q \leqslant C, \tag{2.27}$$

when n is large enough. Inequality (2.27), together with (2.17), further ensures that for large n and all $1 \leqslant k \leqslant K_n$,

$$E\|\hat{R}_n^{-1}(k) - R^{-1}(k)\|^{q/2} \leqslant C\left(\frac{k^2}{N}\right)^{q/4}. \tag{2.28}$$

(Notice that the C's in (2.27) and (2.28) are independent of both n and k.) These uniform moment bounds are important tools for establishing the main results in Section 3. Note that although Papangelou [16, Corollary 2.5] had shown that the term on the left-hand side of (2.21) converges to 0 when K_n is a fixed constant, no rate of convergence has been reported in the existing publications, particularly not in the situation where $K_n \uparrow \infty$. Inequality (2.21) seems to be the first result that provides a rate.

On the other hand, under conditions (K.1) with $\sum_{i=1}^{\infty} |ia_i| < \infty$, and (K.3), Gerencsér [10, Lemmas 3 and 4] showed that for $q \geqslant 1$,

$$E(\|(\hat{R}_n(k) + \delta_n I_k)^{-1} - R^{-1}(k)\|^q) \leqslant C\left(\delta_n + \frac{k}{N^{1/2}} + \frac{k^2}{N\delta_n}\right)^q, \tag{2.29}$$

where $1 \leqslant k \leqslant K_n$, $K_n = o(\sqrt{n})$, I_k denotes the $k \times k$ identity matrix, and δ_n is any positive number. Observe that if one chooses δ_n to satisfy $\delta_n \sim k/\sqrt{N}$, then the left-hand side of (2.29) has the same rate of convergence as that of (2.28). Inequality (2.29) allows us to approximate the inverse of $\hat{R}_n(k) + \delta_n I$ by $R^{-1}(k)$ in an increasing-order setting. This approximation is especially useful in evaluating the MSPE of the ridge regression predictor

$$\hat{x}^*_{n+1}(k) = -\boldsymbol{x}_n'(k)\hat{\mathbf{a}}^*(k), \tag{2.30}$$

where

$$\hat{\mathbf{a}}^*(k) = -(\hat{R}_n(k) + \delta_n I)^{-1}\frac{1}{N}\sum_{j=K_n}^{n-1} \boldsymbol{x}_j(k)x_{j+1}.$$

For more details on $\hat{x}^*_{n+1}(k)$, see Corollary 1, Remark 6 and the discussion preceding them.

3. The MSPE of the least-squares predictor

In this section, our goal is to give an asymptotic expression for the MSPE of the least-squares predictor $\hat{x}_{n+1}(k)$ with $1 \leqslant k \leqslant K_n$. First notice that

$$E(x_{n+1} - \hat{x}_{n+1}(k))^2 = \sigma^2 + E(\boldsymbol{f}(k) + \mathscr{S}_n(k))^2, \tag{3.1}$$

where with

$$e_{j+1,k} = x_{j+1} + \sum_{l=1}^{k} a_l(k) x_{j+1-l},$$

and

$$\mathbf{a}(k) = (a_1(k), \ldots, a_k(k))' = \arg \min_{(c_1,\ldots,c_k)' \in R^k} E\left(x_{k+1} + \sum_{l=1}^{k} c_l x_{k+1-l}\right)^2,$$

$$\boldsymbol{f}(k) = \mathbf{x}_n'(k) \hat{R}_n^{-1}(k) \frac{1}{N} \sum_{j=K_n}^{n-1} \mathbf{x}_j(k) e_{j+1,k};$$

and with $a_i(k) = 0$ for $i > k$,

$$\mathscr{S}_j(k) = \sum_{i=1}^{\infty} (a_i - a_i(k)) x_{j+1-i}.$$

In the following, $\mathbf{a}(k)$ is sometimes viewed as an infinite-dimensional vector with entries $a_i(k)$, $i = 1, 2, \ldots$.

Let us begin with the difference between $N^{-1} \sum_{j=K_n}^{n-1} \mathbf{x}_j(k) e_{j+1,k}$ and $N^{-1} \sum_{j=K_n}^{n-1} \mathbf{x}_j(k) e_{j+1}$. Following Shibata [17, Section 2], define

$$\mathscr{V} = \{\mathbf{l} = (l_1, l_2, \ldots) \in R^{\infty} : \|\mathbf{l}\|_R^2 < \infty\},$$

where

$$\|\mathbf{l}\|_R^2 = \operatorname{var}\left(\sum_{i=1}^{\infty} l_i x_{1-i}\right).$$

Also define an inner product on $\mathscr{V}$ by

$$(\mathbf{s}, \mathbf{t})_R = \operatorname{cov}\left(\sum_{i=1}^{\infty} s_i x_{1-i}, \sum_{i=1}^{\infty} t_i x_{1-i}\right),$$

where $\mathbf{s} = (s_1, s_2, \ldots)$, $\mathbf{t} = (t_1, t_2, \ldots) \in \mathscr{V}$. Then, it is easy to see that $\mathbf{a}(k)$ is the orthogonal projection of $\mathbf{a} = (a_1, a_2, \ldots)'$ on $\mathscr{V}_k \subseteq \mathscr{V}$, where

$$\mathscr{V}_k = \operatorname{Span}\{(1, 0, \ldots)', \ldots, (\underbrace{0, \ldots, 1}_{k}, 0, \ldots)\}.$$

Therefore, for all $\mathbf{w} \in \mathscr{V}_k$,

$$(\mathbf{w}, \mathbf{a} - \mathbf{a}(k))_R = 0. \tag{3.2}$$

It also can be seen that

$$||\mathbf{a}-\mathbf{a}(k)||_R^2 = E(\mathscr{S}_n^2(k)), \tag{3.3}$$

and that

$$||\mathbf{a}-\mathbf{a}(k)||_R^2 \leqslant ||\mathbf{a}-\mathbf{a}^*(k)||_R^2 \leqslant 2\pi f_2 \sum_{i=k+1}^{\infty} a_i^2, \tag{3.4}$$

where $\mathbf{a}^*(k) = (a_1, \ldots, a_k, 0, \ldots)'$, and f_2 can be read from Remark 1.

Lemma 3. *If* (K.1) *holds and* $\sup_{-\infty<t<\infty} E(|e_t|^{2q}) < \infty$ *for some* $q \geqslant 2$, *then for* $1 \leqslant k \leqslant K_n$ *with* $K_n \leqslant n-1$,

$$E\left\|\frac{1}{\sqrt{N}}\sum_{j=K_n}^{n-1} \mathbf{x}_j(k)(e_{j+1,k}-e_{j+1})\right\|^q \leqslant Ck^{q/2}||\mathbf{a}-\mathbf{a}(k)||_R^q. \tag{3.5}$$

Proof. Since

$$\frac{1}{\sqrt{N}}\sum_{j=K_n}^{n-1} \mathbf{x}_j(k)(e_{j+1}-e_{j+1,k}) = \frac{1}{\sqrt{N}}\sum_{j=K_n}^{n-1} \mathbf{x}_j(k)\sum_{i=1}^{\infty} (a_i - a_i(k))x_{j+1-i},$$

by the convexity of x^q, $x>0$,

$$E\left\|\frac{1}{\sqrt{N}}\sum_{j=K_n}^{n-1} \mathbf{x}_j(k)(e_{j+1}-e_{j+1,k})\right\|^q$$
$$\leqslant k^{q/2}k^{-1}\sum_{l=0}^{k-1} E\left\{N^{-q/2}\left|\sum_{j=K_n}^{n-1} x_{j-l}\sum_{i=1}^{\infty} (a_i-a_i(k))x_{j+1-i}\right|^q\right\}. \tag{3.6}$$

By (3.2),

$$E\left(x_{j-l}\sum_{i=1}^{\infty} (a_i-a_i(k))x_{j+1-i}\right) = 0,$$

for $l = 0, 1, \ldots, k-1$. This fact and Theorem 1 yield that the summand (with respect to index l) on the right-hand side of (3.6) is bounded by

$$C\left(\frac{1}{N}\sum_{s=K_n}^{n-1}\sum_{t=K_n}^{n-1} r_{s-t}r_{s-t}^*\right)^{q/2},$$

where

$$r_{s-t}^* = E\left\{\left(\sum_{i=1}^{\infty} (a_i-a_i(k))x_{s+1-i}\right)\left(\sum_{i=1}^{\infty} (a_i-a_i(k))x_{t+1-i}\right)\right\}.$$

Moreover, we have

$$\frac{1}{N}\sum_{s=K_n}^{n-1}\sum_{t=K_n}^{n-1}|r_{s-t}r^*_{s-t}|\leqslant r^*_0\sum_{i=-\infty}^{\infty}|r_i|\leqslant C\|\mathbf{a}-\mathbf{a}(k)\|_R^2,$$

where the second inequality is ensured by $\sum_{j=0}^{\infty}|b_j|<\infty$. As a result, (3.5) follows from these facts and (3.6). □

Remark 4. If $\{e_t, \mathscr{F}_t\}$ is a martingale difference with

$$\sup_{-\infty<t<\infty} E\{|e_t|^{2q}|\mathscr{F}_{t-1}\}<C \quad \text{a.s.,}$$

then (3.5) still holds. Under a Gaussian assumption on $\{e_t\}$, Shibata [17, Lemma 3.1] obtained a result similar to (3.5) for $q=2$. Bhansali [4, Lemma 4.2] and Karagrigoriou [14, Lemma 3.2] also gave similar results in the i.i.d. case for $q=2$. All of these results are special cases of Lemma 3. In addition, if $\{e_t\}$ is a martingale difference sequence, Gerencsér [10, Lemma 5] derived a bound for the left-hand side of (3.5). But it is $N^{q/2}$ times larger than ours.

To obtain the main result of this section, Theorem 3, the following lemma is also needed.

Lemma 4. *If* (K.1) *holds and* $\sup_{-\infty<t<\infty} E\{|e_t|^q\}<\infty$ *for* $q\geqslant 2$, *then for* $1\leqslant k\leqslant K_n$ *with* $K_n\leqslant n-1$,

$$E\left\|\frac{1}{\sqrt{N}}\sum_{j=K_n}^{n-1}\mathbf{x}_j(k)e_{j+1}\right\|^q\leqslant Ck^{q/2}. \tag{3.7}$$

Proof. By an argument similar to that used for showing (3.6),

$$E\left\|\frac{1}{\sqrt{N}}\sum_{j=K_n}^{n-1}\mathbf{x}_j(k)e_{j+1}\right\|^q\leqslant k^{q/2}k^{-1}\sum_{l=0}^{k-1}E\left\{N^{-q/2}\left|\sum_{j=K_n}^{n-1}x_{j-l}e_{j+1}\right|^q\right\}. \tag{3.8}$$

In view of Wei [18, Lemma 2] and the convexity of $x^{q/2}$, $x>0$, the summand (with respect to index l) on the right-hand side of (3.8) is bounded by

$$CE\left(\frac{1}{N}\sum_{j=K_n}^{n-1}x_{j-l}^2\right)^{q/2}\leqslant C\frac{1}{N}\sum_{j=K_n}^{n-1}E\{|x_{j-l}|^q\}.$$

Since $x_t=\sum_{k=0}^{\infty}b_k e_{t-k}$, by Wei [18, Lemma 2] again, $E\{|x_{j-l}|^q\}\leqslant C$, for all integers j and l. These results and (3.8) yield the desired result. □

An asymptotic expression for the MSPE of $\hat{x}_{n+1}(k)$, which holds uniformly for all $1\leqslant k\leqslant K_n$, is given in the following theorem.

Theorem 3. *Assume* (K.1) *with* $\sum_{i=1}^{\infty} i^{1/2}|a_i| < \infty$, (K.2), (K.3), *and* $K_n^{2+\delta_1} = O(n)$ *for some* $\delta_1 > 0$. *Then,*

$$\lim_{n\to\infty} \max_{1\leqslant k\leqslant K_n} \left| \frac{E(x_{n+1} - \hat{x}_{n+1}(k))^2 - \sigma^2}{L_n(k)} - 1 \right| = 0, \tag{3.9}$$

where $L_n(k) = (k/N)\sigma^2 + ||\mathbf{a} - \mathbf{a}(k)||_R^2$.

When model AR(k) is considered, $L_n(k)$ can be viewed as a measure of the model's quality of prediction, which is the sum of model complexity, $(k/N)\sigma^2$, and goodness of fit, $||\mathbf{a} - \mathbf{a}(k)||_R^2$. Moreover, since (3.9) shows that for large n, the *second-order* MSPE of $\hat{x}_{n+1}(k)$, $E(x_{n+1} - \hat{x}_{n+1}(k))^2 - \sigma^2$, can be *uniformly* approximated by $L_n(k)$, a model with order k_n^* satisfying $L_n(k_n^*) = \min_{1\leqslant k\leqslant K_n} L_n(k)$ can be viewed as the best choice among models AR(1), ..., AR(K_n), from a prediction point of view. For example, if for some $0 < c_1 \leqslant c_2 < \infty$ and $\beta > 0$, the AR coefficients satisfy

$$c_1 e^{-\beta k} \leqslant \sum_{i\geqslant k} a_i^2 \leqslant c_2 e^{-\beta k} \tag{3.10}$$

(a condition fulfilled by any causal and invertible ARMA(p,q) process with $p \geqslant 0$ and $q > 0$), then, by some algebraic manipulations, $k_n^* \sim (1/\beta)\log n$. In addition, if for some $0 < c_3 \leqslant c_4 < \infty$ and $\beta > 0$, the coefficients satisfy

$$c_3 k^{-\beta} \leqslant \sum_{i\geqslant k} a_i^2 \leqslant c_4 k^{-\beta} \tag{3.11}$$

(see [17, p. 162] for a similar example), then it is also not difficult to show that for some $0 < c_3' \leqslant c_4' < \infty$, $c_3' N^{1/(\beta+1)} < k_n^* \leqslant c_4' N^{1/(\beta+1)}$ for large n.

Proof of Theorem 3. By (3.1),

$$\max_{1\leqslant k\leqslant K_n} \left| \frac{E(x_{n+1} - \hat{x}_{n+1}(k))^2 - \sigma^2}{L_n(k)} - 1 \right| = \max_{1\leqslant k\leqslant K_n} \left| \frac{E(\boldsymbol{f}(k) + \mathscr{S}_n(k))^2}{L_n(k)} - 1 \right|. \tag{3.12}$$

In view of (3.12), our proof is divided into four steps.

Step 1: Prove that

$$\lim_{n\to\infty} \max_{1\leqslant k\leqslant K_n} E\left\{ \left| \sqrt{\frac{N}{k}} (\boldsymbol{f}(k) - \boldsymbol{f}_0(k)) \right|^q \right\} = 0, \tag{3.13}$$

for all $q > 0$, where

$$\boldsymbol{f}_0(k) = \boldsymbol{x}_n^{*'}(k) \hat{R}_n^{\circ^{-1}}(k) N^{-1} \sum_{j=K_n}^{n-\sqrt{n}-1} \boldsymbol{x}_j(k) e_{j+1,k},$$

with

$$\boldsymbol{x}_n^*(k) = (x_n^*, \ldots, x_{n-k+1}^*)' = \left(\sum_{j=0}^{\sqrt{n}/2-K_n} b_j e_{n-j}, \ldots, \sum_{j=0}^{\sqrt{n}/2-K_n} b_j e_{n-k+1-j} \right)',$$

$$\hat{R}_n{}^{\circ}(k) = N^{-1} \sum_{j=K_n}^{n-\sqrt{n}-1} \boldsymbol{x}_j(k)\boldsymbol{x}_j'(k),$$

and $\hat{R}_n{}^{\circ^{-1}}(k)$ is defined by the same conventions as $\hat{R}_n^{-1}(k)$.

Without loss of generality, we assume that $q \geqslant 2/3$, since the result for $q < 2/3$ can be obtained from the result for $q \geqslant 2/3$ and Jensen's inequality. Observe that

$$\begin{aligned}
&|\sqrt{N/k}(\boldsymbol{f}(k) - \boldsymbol{f}_0(k))| \\
&\leqslant \Bigg| k^{-1/2}(\boldsymbol{x}_n(k) - \boldsymbol{x}_n^*(k))'\hat{R}_n^{-1}(k)N^{-1/2} \sum_{j=K_n}^{n-1} \boldsymbol{x}_j(k)e_{j+1,k} \\
&\quad + k^{-1/2}\boldsymbol{x}_n^*(k)(\hat{R}_n^{-1}(k) - \hat{R}_n{}^{\circ^{-1}}(k))N^{-1/2} \sum_{j=K_n}^{n-1} \boldsymbol{x}_j(k)e_{j+1,k} \\
&\quad + k^{-1/2}\boldsymbol{x}_n^*(k)\hat{R}_n{}^{\circ^{-1}}(k)N^{-1/2} \sum_{j=n-\sqrt{n}}^{n-1} \boldsymbol{x}_j(k)e_{j+1,k} \Bigg|. \qquad (3.14)
\end{aligned}$$

By Wei [18, Lemma 2] and the convexity of the function $x^{3q/2},\ x > 0$,

$$\max_{1 \leqslant k \leqslant K_n} E(k^{-1/2}||\boldsymbol{x}_n(k) - \boldsymbol{x}_n^*(k)||)^{3q} \leqslant C \left(\sum_{j=\sqrt{n}/2-K_n+1}^{\infty} b_j^2 \right)^{3q/2}. \qquad (3.15)$$

Similarly,

$$\max_{1 \leqslant k \leqslant K_n} E(k^{-1/2}||\boldsymbol{x}_n^*(k)||)^{3q} \leqslant C. \qquad (3.16)$$

Reasoning as for (2.27) and (2.28), we have for large n,

$$\max_{1 \leqslant k \leqslant K_n} E||\hat{R}_n{}^{\circ^{-1}}(k)||^{3q} \leqslant C$$

and

$$E||\hat{R}_n{}^{\circ^{-1}}(k) - R^{-1}(k)||^{3q} \leqslant C(k^2/N)^{3q/2} \qquad (3.17)$$

for all $1 \leqslant k \leqslant K_n$. By (2.27), (3.17), and arguments similar to those used for verifying Lemma 2 and Theorem 2, we obtain for large n and all $1 \leqslant k \leqslant K_n$,

$$
\begin{aligned}
& E||\hat{R}_n^{-1}(k) - \hat{R}_n^{\circ^{-1}}(k)||^{3q} \\
& \leqslant \left(E||\hat{R}_n^{-1}(k)||^{9q} E||\hat{R}_n^{\circ^{-1}}(k)||^{9q} E\left|\left|\frac{1}{N}\sum_{j=n-\sqrt{n}}^{n-1} \boldsymbol{x}_j(k)\boldsymbol{x}_j'(k)\right|\right|^{9q}\right)^{1/3} \\
& \leqslant C(k^{3q}N^{-9q/4} + N^{-3q/2}).
\end{aligned} \tag{3.18}
$$

Furthermore, by Lemmas 3 and 4 for all $1 \leqslant k \leqslant K_n$,

$$
\begin{aligned}
& E\left|\left| N^{-1/2} \sum_{j=K_n}^{n-1} \boldsymbol{x}_j(k) e_{j+1,k} \right|\right|^{3q} \\
& \leqslant C\left(E\left|\left| N^{-1/2} \sum_{j=K_n}^{n-1} \boldsymbol{x}_j(k)(e_{j+1,k} - e_{j+1}) \right|\right|^{3q} \right. \\
& \qquad \left. + \; E\left|\left| N^{-1/2} \sum_{j=K_n}^{n-1} \boldsymbol{x}_j(k) e_{j+1} \right|\right|^{3q} \right) \\
& \leqslant C(k^{3q/2}||\mathbf{a} - \mathbf{a}(k)||_R^{3q} + k^{3q/2}).
\end{aligned} \tag{3.19}
$$

Similarly, for all $1 \leqslant k \leqslant K_n$,

$$
\begin{aligned}
& E\left|\left| N^{-1/2} \sum_{j=n-\sqrt{n}}^{n-1} \boldsymbol{x}_j(k) e_{j+1,k} \right|\right|^{3q} \\
& \leqslant CN^{-3q/4}(k^{3q/2}||\mathbf{a} - \mathbf{a}(k)||_R^{3q} + k^{3q/2}).
\end{aligned} \tag{3.20}
$$

Consequently, (3.13) follows from Hölder's inequality, (3.4), (3.14)–(3.20), and the fact that $\sum_{i=1}^{\infty} |i^{1/2} b_i| < \infty$, which is ensured by $\sum_{i=1}^{\infty} |i^{1/2} a_i| < \infty$ (see [6, Theorem 3.8.4]).

Step 2: Prove that

$$
\lim_{n \to \infty} \max_{1 \leqslant k \leqslant K_n} E|L_n^{-1/2}(k)(\boldsymbol{f}_0(k) - \boldsymbol{f}_1(k))|^q = 0, \tag{3.21}
$$

for all $q > 0$, where $\boldsymbol{f}_1(k) = \boldsymbol{x}_n^{*'}(k) R^{-1}(k) N^{-1} \sum_{j=K_n}^{n-\sqrt{n}-1} \boldsymbol{x}_j(k) e_{j+1}$.

Assume $q \geqslant 2$. It is easy to see that for all $1 \leqslant k \leqslant K_n$,

$$\begin{aligned} & E\{(\sqrt{N/k}|\boldsymbol{f}_0(k) - \boldsymbol{f}_1(k)|)^q\} \\ & \leqslant C\left(E\left| k^{-\frac{1}{2}} \boldsymbol{x}_n^{*'}(k)(\hat{R}_n^{\circ^{-1}}(k) - R^{-1}(k))N^{-\frac{1}{2}} \sum_{j=K_n}^{n-\sqrt{n}-1} \boldsymbol{x}_j(k) e_{j+1,k} \right|^q \right. \\ & \quad \left. + E\left| k^{-\frac{1}{2}} \boldsymbol{x}_n^{*'}(k) R^{-1}(k) N^{-\frac{1}{2}} \sum_{j=K_n}^{n-\sqrt{n}-1} \boldsymbol{x}_j(k)(e_{j+1,k} - e_{j+1}) \right|^q \right). \end{aligned} \tag{3.22}$$

Since $\boldsymbol{x}_n^*(k)$ is independent of $(e_{n-\sqrt{n}}, e_{n-\sqrt{n}-1}, \ldots)'$, the right-hand side of (3.22) can be rewritten as

$$CE\left(\int_{\boldsymbol{x} \in R^k} |\boldsymbol{x}'\boldsymbol{u}_1|^q dF_{\boldsymbol{x}_n^*(k)}(\boldsymbol{x}) + \int_{\boldsymbol{x} \in R^k} |\boldsymbol{x}'\boldsymbol{u}_2|^q dF_{\boldsymbol{x}_n^*(k)}(\boldsymbol{x}) \right), \tag{3.23}$$

where $F_{\boldsymbol{x}_n^*(k)}(\cdot)$ denotes the joint distribution function of $\boldsymbol{x}_n^*(k)$,

$$\boldsymbol{u}_1 = k^{-\frac{1}{2}} N^{-\frac{1}{2}} (\hat{R}_n^{\circ^{-1}}(k) - R^{-1}(k)) \sum_{j=K_n}^{n-\sqrt{n}-1} \boldsymbol{x}_j(k) e_{j+1,k}$$

and

$$\boldsymbol{u}_2 = k^{-\frac{1}{2}} N^{-\frac{1}{2}} R^{-1}(k) \sum_{j=K_n}^{n-\sqrt{n}-1} \boldsymbol{x}_j(k)(e_{j+1,k} - e_{j+1}).$$

Since

$$\int_{\boldsymbol{x} \in R^k} |\boldsymbol{x}'\boldsymbol{u}_i|^q dF_{\boldsymbol{x}_n^*(k)}(\boldsymbol{x}), \quad i = 1, 2$$

are qth moments of linear combinations in $\{e_i, \ -\infty < i \leqslant n\}$, by Wei [18, Lemma 2], we obtain that for $i = 1$ and 2,

$$\int_{\boldsymbol{x} \in R^k} |\boldsymbol{x}'\boldsymbol{u}_i|^q dF_{\boldsymbol{x}_n^*(k)}(\boldsymbol{x}) \leqslant C|\boldsymbol{u}_i' R^*(k) \boldsymbol{u}_i|^{q/2}, \tag{3.24}$$

where $R^*(k) = \mathrm{E}(\boldsymbol{x}_1^*(k)\boldsymbol{x}_1^{*'}(k))$. Simple algebraic manipulations yield

$$\begin{aligned} & E|\boldsymbol{u}_1' R^*(k) \boldsymbol{u}_1|^{\frac{q}{2}} \\ & \leqslant E\left(||R^*(k)||^{\frac{q}{2}} ||\hat{R}_n^{\circ^{-1}}(k) - R^{-1}(k)||^q \left|\left| k^{-\frac{1}{2}} N^{-\frac{1}{2}} \sum_{j=K_n}^{n-\sqrt{n}-1} \boldsymbol{x}_j(k) e_{j+1,k} \right|\right|^q \right) \end{aligned}$$

and

$$\begin{aligned} & E|\boldsymbol{u}_2' R^*(k) \boldsymbol{u}_2|^{\frac{q}{2}} \\ & \leqslant E\left(||R^*(k)||^{\frac{q}{2}} ||R^{-1}(k)||^q \left|\left| k^{-\frac{1}{2}} N^{-\frac{1}{2}} \sum_{j=K_n}^{n-\sqrt{n}-1} \boldsymbol{x}_j(k)(e_{j+1,k} - e_{j+1}) \right|\right|^q \right). \end{aligned}$$

Since $\sup_{k\geqslant 1}||R^*(k)||\leqslant C$ is ensured by the absolute summability of b_i's, this fact, (3.17), and Lemmas 3 and 4 together imply that for large n and all $1\leqslant k\leqslant K_n$,

$$E|\boldsymbol{u}_1' R^*(k)\boldsymbol{u}_1|^{q/2}\leqslant C(k^2/N)^{q/2} \tag{3.25}$$

and

$$E|\boldsymbol{u}_2' R^*(k)\boldsymbol{u}_2|^{q/2}\leqslant C||\mathbf{a}-\mathbf{a}(k)||_R^q. \tag{3.26}$$

In view of (3.22)–(3.26), one obtains that for large n and all $1\leqslant k\leqslant K_n$,

$$\begin{aligned} E\left\{\frac{|\boldsymbol{f}_0(k)-\boldsymbol{f}_1(k)|^2}{L_n(k)}\right\}^{q/2} &\leqslant C\left\{\left(\frac{k^3}{N^2L_n(k)}\right)^{q/2}+\left(\frac{k||\mathbf{a}-\mathbf{a}(k)||_R^2}{NL_n(k)}\right)^{q/2}\right\}\\ &\leqslant C\left(\frac{K_n^2}{N}\right)^{q/2}. \end{aligned}$$

This yields (3.21) with $q\geqslant 2$, and hence for all $q>0$ by the Jensen inequality.

Step 3: Prove that

$$\lim_{n\to\infty}\max_{1\leqslant k\leqslant K_n}\left|E\left(\frac{N}{k\sigma^2}\boldsymbol{f}_1^2(k)\right)-1\right|=0. \tag{3.27}$$

Equality (3.27) follows from observing that

$$E\left(\frac{N}{k}\boldsymbol{f}_1^2(k)\right)=\mathrm{tr}(R^{-1}(k)R^*(k)k^{-1})(N-\sqrt{n})N^{-1}\sigma^2,$$

and

$$\begin{aligned} \max_{1\leqslant k\leqslant K_n}|\mathrm{tr}(R^{-1}(k)R^*(k)k^{-1})-1| &= \max_{1\leqslant k\leqslant K_n}|\mathrm{tr}(R^{-1}(k)(R^*(k)-R(k))k^{-1})|\\ &\leqslant \max_{1\leqslant k\leqslant K_n}||R^{-1/2}(k)||^2\max_{1\leqslant k\leqslant K_n}||R(k)-R^*(k)||\\ &\leqslant \max_{1\leqslant k\leqslant K_n}||R^{-1/2}(k)||^2\max_{1\leqslant k\leqslant K_n}k\sum_{l=\sqrt{n}/2-2K_n+2}^{\infty}b_l^2\\ &=o(1), \end{aligned}$$

where the last equality is ensured by (K.1), the condition on K_n, and $\sum_{i=1}^{\infty}|i^{1/2}a_i|<\infty$.

Now drawing a conclusion from (3.13), (3.21) and (3.27), we have for all $q>0$,

$$\lim_{n\to\infty}\max_{1\leqslant k\leqslant K_n}E(L_n^{-q/2}(k)\,|\boldsymbol{f}(k)-\boldsymbol{f}_1(k)|^q)=0 \tag{3.28}$$

and

$$\lim_{n\to\infty}\max_{1\leqslant k\leqslant K_n}\left|E\left\{L_n^{-1}(k)\left(\boldsymbol{f}^2(k)-\frac{k}{N}\sigma^2\right)\right\}\right|=0. \tag{3.29}$$

The final step deals with the cross-product term.

Step 4: Prove that

$$\lim_{n\to\infty}\max_{1\leqslant k\leqslant K_n}|E\{L_n^{-1}(k)\boldsymbol{f}(k)\mathscr{S}_n(k)\}|=0. \tag{3.30}$$

Let

$$\mathscr{S}_{1,n}(k) = \sum_{i=1}^{\sqrt{n}/2} (a_i - a_i(k))x^{**}_{n+1-i},$$

where $x^{**}_{n+1-i} = \sum_{j=0}^{\sqrt{n}/2} b_j e_{n+1-i-j}$. Since $\boldsymbol{x}^*_n(k)$ is independent of $(\mathscr{S}_n(k) - \mathscr{S}_{1,n}(k), \sum_{j=K_n}^{n-\sqrt{n}-1} \boldsymbol{x}_j(k)e_{j+1})$, and $\sum_{j=K_n}^{n-\sqrt{n}-1} \boldsymbol{x}_j(k)e_{j+1}$ is independent of $(\mathscr{S}_{1,n}(k), \boldsymbol{x}^*_n(k))$, one obtains

$$\begin{aligned}
&|E(\boldsymbol{f}(k)\mathscr{S}_n(k)L_n^{-1}(k))| \\
&\quad = |E[\{(\boldsymbol{f}(k) - \boldsymbol{f}_1(k))\mathscr{S}_n(k) + \boldsymbol{f}_1(k)(\mathscr{S}_n(k) - \mathscr{S}_{1,n}(k)) + \boldsymbol{f}_1(k)\mathscr{S}_{1,n}(k)\}L_n^{-1}(k)]| \\
&\quad = |E\{(\boldsymbol{f}(k) - \boldsymbol{f}_1(k))\mathscr{S}(k)L_n^{-1}(k)\}|.
\end{aligned}$$

By the Cauchy–Schwarz inequality and (3.28),

$$\begin{aligned}
&\max_{1\leqslant k\leqslant K_n} |E\{(\boldsymbol{f}(k) - \boldsymbol{f}_1(k))\mathscr{S}_n(k)L_n^{-1}(k)\}| \\
&\quad \leqslant \left[\max_{1\leqslant k\leqslant K_n} E\{(\boldsymbol{f}(k) - \boldsymbol{f}_1(k))^2 L_n^{-1}(k)\} \max_{1\leqslant k\leqslant K_n} E(\mathscr{S}_n^2(k)L_n^{-1}(k))\right]^{1/2} = o(1).
\end{aligned}$$

Therefore, (3.30) follows. Now (3.9) is ensured by (3.29), (3.30), and (3.3). □

Remark 5. (A continuation of Remark 3). As observed in the proof of Theorem 3, (2.20), (2.21) and their applications play important roles in obtaining (3.9). However, if (2.15) is used in the proof of Theorem 2 instead of (2.16), then to obtain (2.20) and (2.21), the divergence rate of the maximal order K_n must be confined to $o(\log n)$, as shown in Remark 3. This limitation ultimately excludes the optimal order, k^*_n, in the examples given previously. Therefore, if our analysis had started from (2.15), it would not be clear whether (3.9) holds for these $\hat{x}_{n+1}(k^*_n)$'s.

4. Some comparisons

In this section, we first compare the MSPEs of the least-squares predictors for same- and independent-realizations. For independent-realization predictions, Shibata [17, Proposition 3.2], assuming (K.1) with $a_i \neq 0$ for infinitely many i, $K_n = o(n^{1/2})$, and Gaussian noise, showed that

$$\max_{1\leqslant k\leqslant K_n} \left| \frac{E\{(y_{n+1} - \hat{y}_{n+1}(k))^2 | x_1, \ldots, x_n\} - \sigma^2}{L_n(k)} - 1 \right| = o_p(1), \tag{4.1}$$

where $y_1, \ldots, y_n$ are observations from an independent replicate of $\{x_t\}$, $\hat{y}_{n+1}(k) = -\boldsymbol{y}'_n(k)\hat{\mathbf{a}}(k), \hat{\mathbf{a}}(k)$, defined in (1.5), is the least-squares estimator obtained from $x_1, \ldots, x_n$; and $\boldsymbol{y}_n(k) = (y_n, \ldots, y_{n-k+1})'$. As observed in (4.1), $L_n(k)$ can be used for uniformly approximating the *conditional* MSPE of $\hat{y}_{n+1}(k)$. However, since (3.9) focuses on the *unconditional* MSPE in a same-realization setting, for the purpose of comparison, we now provide an unconditional version of (4.1).

Theorem 4. *Assume that the assumptions of Theorem* 3 *hold. Then*,

$$\lim_{n\to\infty} \max_{1\leqslant k\leqslant K_n} \left| \frac{E\{(y_{n+1}-\hat{y}_{n+1}(k))^2\}-\sigma^2}{L_n(k)} - 1 \right| = 0. \tag{4.2}$$

Proof. By Shibata [17, (2.5)],

$$E\{(y_{n+1}-\hat{y}_{n+1}(k))^2|x_1,\ldots,x_n\}-\sigma^2 = ||\hat{\mathbf{a}}(k)-\mathbf{a}(k)||_R^2 + ||\mathbf{a}(k)-\mathbf{a}||_R^2, \tag{4.3}$$

where $\hat{\mathbf{a}}(k)$ and $\mathbf{a}(k)$ are now viewed as infinite-dimensional vectors with undefined entries set to 0. Therefore,

$$E\{(y_{n+1}-\hat{y}_{n+1}(k))^2\}-\sigma^2 = E(||\hat{\mathbf{a}}(k)-\mathbf{a}(k)||_R^2) + ||\mathbf{a}(k)-\mathbf{a}||_R^2.$$

Following an argument similar to that used for verifying Theorem 3, it can be shown that

$$\max_{1\leqslant k\leqslant K_n} \left| \frac{E(||\hat{\mathbf{a}}(k)-\mathbf{a}(k)||_R^2) - \frac{k\sigma^2}{N}}{L_n(k)} \right| = o(1),$$

which yields the desired result. □

From (3.9) and (4.2), it can be seen that both types of second-order MSPEs can be uniformly approximated by the same function, $L_n(k)$. This result suggests that an estimated AR model that has good ability to forecast the future of an independent replicate will also perform well in predicting the future of the observed time series. This further leads us to conjecture that the second-order MSPE of the predictor with order selected by AIC will ultimately achieve the minimal $L_n(k)$ value in the same-realization setting (this property is referred to as the asymptotic efficiency), because Shibata [17] had shown that AIC possesses a similar property for independent-realization predictions. Through clarifying the dependence structures among the model-order selectors, the estimated parameters, and future observations in same-realization and increasing-order settings, we provide the first theoretical verification that AIC is asymptotically efficient for same-realization predictions in a companion paper [13].

It is worth noting that the asymptotic equivalence between second-order MSPEs in same- and independent-realization settings, as shown in Theorems 3 and 4, should not be taken for granted. To see this, Ing [11] recently showed that if the underlying process is a random walk model, and the assumed model is correctly specified, then

$$\lim_{n\to\infty} \frac{E(x_{n+1}-\hat{x}_{n+1}(1))^2-\sigma^2}{\frac{\sigma^2}{n}} = 2$$

and

$$\lim_{n\to\infty} \frac{E(y_{n+1}-\hat{y}_{n+1}(1))^2-\sigma^2}{\frac{\sigma^2}{n}} = 13.2859.$$

Hence,

$$\lim_{n\to\infty} \frac{E(x_{n+1}-\hat{x}_{n+1}(1))^2-\sigma^2}{E(y_{n+1}-\hat{y}_{n+1}(1))^2-\sigma^2} = \frac{2}{13.2859}.$$

As observed, the second-order MSPE for same-realization predictions is much smaller than that for independent-realization predictions. Therefore, the equivalence just mentioned does not hold in this example.

Under stationary AR processes, Kunitomo and Yamamoto [15, pp. 946–947] also considered a comparison between these MSPEs in the situation where the assumed fixed-order AR model is underspecified. They showed that the difference between the terms of order $1/n$ in two types of MSPEs can be substantial, but neither of them is uniformly better. (Note that their conclusion does not contradict those obtained from Theorems 3 and 4, because their main concern is with the terms of order $1/n$, but the second-order MSPEs are of order $O(1)$ in the underspecified and fixed-order case (see (3.9)).) The above results show that the difference between the MSPEs in two types of forecasting settings should be carefully examined in each different situation, and that it can be erroneous to directly assume that the results for same-realization predictions will be the same as those for corresponding independent cases without theoretical justification.

When the smoothness condition on e_t, (K.2), is removed from our analysis, one may encounter the possibly ill-conditioned matrix, $\hat{R}_n(k)$, in dealing with the moment properties of the least-squares predictor, $\hat{x}_{n+1}(k)$. This problem becomes more serious in increasing-order settings. To overcome this difficulty, the ridge regression predictor, $\hat{x}^*_{n+1}(k)$ (see (2.30)), is a possible remedy. In the following, we investigate the performance of $\hat{x}^*_{n+1}(k)$ in increasing-order settings.

Assume (K.1) with $\sum_{i=1}^{\infty}|ia_i|<\infty$, (K.3), $K_n=o(n^{1/7})$, $\delta_n \sim n^{-3/4}k^{5/4}$, and

$$k^2\|\boldsymbol{a}-\boldsymbol{a}(k)\|_R^2 = O(n^{-1}). \tag{4.4}$$

Then, Gerencsér [10, Theorem 2] proved that

$$E(x_{n+1}-\hat{x}^*_{n+1}(k))^2-\sigma^2 = L_n(k)+o(L_n(k)). \tag{4.5}$$

First observe that the expression for the second-order MSPE of $\hat{x}^*_{n+1}(k)$ in (4.5) is the same as that of $\hat{x}_{n+1}(k)$ in (3.9). Therefore, k_n^* (see Section 3) is the common optimal order for these two predictors. Although (4.5) holds without (K.2), condition (4.4) is too stringent. To see this, consider the ARMA case (3.10). In this case, $k_n^* \sim (1/\beta)\log n$, and hence $nk_n^{*^2}\|\boldsymbol{a}-\boldsymbol{a}(k_n^*)\|_R^2 \to \infty$. Because (4.4) is violated, with Gerencsér's approach, it is not clear whether (4.5) holds for $\hat{x}^*_{n+1}(k_n^*)$. The same difficulty also arises in the algebraic-decay case (3.11). We also note that Gerencsér's expression for the MSPE of $\hat{x}^*_{n+1}(k)$ does not hold uniformly for all $1 \leqslant k \leqslant K_n$.

To remove these difficulties, we can use (2.29) instead of (2.28) and follow the same line of argument as that used for verifying Theorem 3.1 to obtain the following result.

Corollary 1. *Assume* (K.1) *with* $\sum_{i=1}^{\infty} |i^{1/2} a_i| < \infty$, *and* (K.3). *Then, for* $1 \leqslant k \leqslant K_n$, $K_n^3 = o(n)$, *and* $\delta_n \sim n^{-3/4} k^{5/4}$,

$$\lim_{n\to\infty} \max_{1\leqslant k\leqslant K_n} \left| \frac{E(x_{n+1} - \hat{x}^{*}_{n+1}(k))^2 - \sigma^2}{L_n(k)} - 1 \right| = 0. \tag{4.6}$$

Remark 6. Since this corollary does not need condition (4.4), which imposes a strong connection between k and $||\boldsymbol{a} - \boldsymbol{a}(k)||_R^2$, the difficulties mentioned above are avoided. Moreover, the asymptotic equivalence between the MSPEs of least square and ridge regression predictors is also established by Theorem 3 and Corollary 1 under their rather mild conditions.

Before leaving this section, we consider multi-step-ahead generalizations of (3.9). First notice that from model (1.1), x_{t+h} for $h \geqslant 1$ can be expressed as

$$-\sum_{i=1}^{\infty} a_{i,h} x_{t-i+1} + \sum_{i=0}^{h-1} b_i e_{t+h-i},$$

where

$$(a_{1,h}, \ldots)' = \arg \min_{(c_1,\ldots)' \in R^{\infty}} E\left(x_{t+h} + \sum_{i=1}^{\infty} c_i x_{t-i+1} \right)^2.$$

Also define

$$(a_{1,h}(k), \ldots, a_{k,h}(k))' = \arg \min_{(c_1,\ldots,c_k)' \in R^{k}} E\left(x_{t+h} + \sum_{i=1}^{k} c_i x_{t-i+1} \right)^2.$$

Under the assumptions of Theorem 3, a multi-step-ahead generalization of (3.9) is given as follows:

$$\lim_{n\to\infty} \max_{1\leqslant k\leqslant K_n} \left| \frac{E(x_{n+h} - \hat{x}_{n+h}(k))^2 - \sigma_h^2}{L_n^h(k)} - 1 \right| = 0, \tag{4.7}$$

where $h \geqslant 1$, $\sigma_h^2 = \sigma^2 \sum_{i=0}^{h-1} b_i^2$, $\hat{x}_{n+h}(k) = -\mathbf{x}_n'(k)\hat{\mathbf{a}}_h(k)$ with

$$-\hat{\mathbf{a}}_h(k) = \left(\sum_{j=K_n}^{n-h} \mathbf{x}_j(k)\mathbf{x}_j'(k) \right)^{-1} \sum_{j=K_n}^{n-h} \mathbf{x}_j(k) x_{j+h},$$

and

$$L_n^h(k) = \frac{\sigma^2 tr(R^{-1}(k) G_h(k))}{N_h} + ||\mathbf{a}_h - \mathbf{a}_h(k)||_R^2$$

with $N_h = n - h - K_n + 1$,

$$G_h(k) = E\left\{\left(\sum_{i=0}^{h-1} b_{h-1-i}\mathbf{x}_{n-i}(k)\right)\left(\sum_{i=0}^{h-1} b_{h-1-i}\mathbf{x}'_{n-i}(k)\right)\right\},$$

$\mathbf{a}_h = (a_{1,h}, \ldots)'$, and $\mathbf{a}_h(k) = (a_{1,h}(k), \ldots, a_{k,h}(k), 0, \ldots)$. Since it is straightforward to verify (4.7) through an argument similar to that used for showing Theorem 3, we omit the details in order to save space.

For independent-realization predictions, Bhansali [4, Proposition 4.1], assuming (K.1) with $a_i \neq 0$ for infinitely many i and $\{e_t\}$ being a sequence of i.i.d. random variables, $K_n = o(n^{1/2})$, and $E(|e_1|^{16}) < \infty$, showed that

$$\max_{1\leqslant k\leqslant K_n}\left|\frac{E\{(y_{n+h} - \hat{y}_{n+h}(k))^2|x_1, \ldots, x_n\} - \sigma_h^2}{L_{0,n}^h(k)} - 1\right| = o_p(1), \tag{4.8}$$

where $h \geqslant 1$, $\hat{y}_{n+h}(k) = -\boldsymbol{y}'_n(k)\hat{\mathbf{a}}_h(k)$
and

$$L_{0,n}^h(k) = \frac{\sigma_h^2 k}{N_h} + ||\mathbf{a}_h - \mathbf{a}_h(k)||_R^2.$$

Since (K.1) ensures that

$$\max_{1\leqslant k\leqslant K_n} |L_n^h(k) - L_{0,n}^h(k)| \leqslant \frac{C}{N_h},$$

one has

$$\max_{1\leqslant k\leqslant K_n}\left|\frac{L_n^h(k) - L_{0,n}^h(k)}{L_{0,n}^h(k)}\right| \leqslant \frac{C}{N_h L_{0,n}^h(k_{n,h}^*)} = o(1), \tag{4.9}$$

provided $a_i \neq 0$ for infinitely many i. Here,

$$L_{0,n}^h(k_{n,h}^*) = \min_{1\leqslant k\leqslant K_n} L_{0,n}^h(k),$$

and the equality in (4.9) follows from $k_{n,h}^* \to \infty$ as $n \to \infty$ (see [4, p. 584]).

When $\{x_t\}$ is truly an AR (∞) process, by (4.9) and some algebraic manipulations, (4.7) still holds with $L_n^h(k)$ replaced by $L_{0,n}^h(k)$, namely,

$$\lim_{n\to\infty}\max_{1\leqslant k\leqslant K_n}\left|\frac{E(x_{n+h} - \hat{x}_{n+h}(k))^2 - \sigma_h^2}{L_{0,n}^h(k)} - 1\right| = 0. \tag{4.10}$$

On the other hand, if $\{x_t\}$ is an AR (p) process with $1 \leqslant p < \infty$, then for all $n \geqslant 1$,

$$\frac{L_n^h(p) - L_{0,n}^h(p)}{L_{0,n}^h(p)} = \frac{\sigma^2 \mathrm{tr}(R^{-1}(p)G_h(p)) - \sigma_h^2 p}{\sigma_h^2 p}, \tag{4.11}$$

where the equality follows from $||\mathbf{a}_h - \mathbf{a}_h(p)||_R^2 = 0$. Moreover, by Eq. (14) of [12], one has for $h = 2$ and $a_1 \neq 0$,

$$\frac{\sigma^2 \mathrm{tr}(R^{-1}(p)G_h(p)) - \sigma_h^2 p}{\sigma_h^2 p} = \frac{2a_1^2}{(1+a_1^2)p} > 0,$$

which, together with (4.7) and (4.11), yields that (4.10) with $h > 1$ is no longer true for finite-oder AR processes. Consequently, $L_n^h(k)$ is a better approximation of $E(x_{n+h} - \hat{x}_{n+h}(k))^2 - \sigma_h^2$ than $L_{0,n}^h(k)$ when $\{x_t\}$ has a possibly finite order.

References

[1] H. Akaike, Statistical predictor identification, Ann. Inst. Statist. Math. 22 (1970) 203–217.
[2] R.J. Bhansali, Linear prediction by autoregressive model fitting in the time domain, Ann. Statist. 6 (1978) 224–231.
[3] R.J. Bhansali, Effects of not knowing the order of an autoregressive process on the mean squared error of prediction—I, J. Amer. Statist. Assoc. 76 (1981) 588–597.
[4] R.J. Bhansali, Asymptotically efficient autoregressive model selection for multistep prediction, Ann. Inst. Statist. Math. 48 (1996) 577–602.
[5] R.J. Bhansali, F. Papangelou, Convergence of moments of least squares estimators for the coefficients of an autoregressive process of unknown order, Ann. Statist. 19 (1991) 1155–1162.
[6] P.J. Brillinger, Time Series: Data Analysis and Theory, Holt, New York, 1975.
[7] D.F. Findley, C.Z. Wei, Moment bounds for deriving time series CLT's and model selection procedures, Statist. Sinica 3 (1993) 453–470.
[8] D.F. Findley, C.Z. Wei, AIC, overfitting principles, and boundedness of moments of inverse matrices for vector autoregressive models fit to weakly stationary time series, J. Multivariate Anal. 83 (2002) 415–450.
[9] W.A. Fuller, D.P. Hasza, Properties of predictors for autoregressive time series, J. Amer. Statist. Assoc. 76 (1981) 155–161.
[10] L. Gerencsér, AR (∞) estimation and nonparametric stochastic complexity, IEEE Trans. Inform. Theory 38 (1992) 1768–1778.
[11] C.K. Ing, A note on mean-squared prediction errors of the least squares predictors in random walk models, J. Time Ser. Anal. 22 (2001) 711–724.
[12] C.K. Ing, Multistep prediction in autoregressive processes, Economet. Theory 19 (2003) 252–277.
[13] C.K. Ing, C.Z. Wei, Order selection for same-realization predictions in autoregressive processes, Technical report C-00-09, Institute of Statistical Science, Academia Sinica, Taipei, Taiwan, ROC, 2000.
[14] A. Karagrigoriou, Asymptotic efficiency of the order selection of a nongaussian AR process, Statist. Sinica 7 (1997) 407–423.
[15] N. Kunitomo, T. Yamamoto, Properties of predictors in misspecified autoregressive time series, J. Amer. Statist. Assoc. 80 (1985) 941–950.
[16] F. Papangelou, On a distributional bound arising in autoregressive model fitting, J. Appl. Probab. 31 (1994) 401–408.
[17] R. Shibata, Asymptotically efficient selection of the order of the model for estimating parameters of a linear process, Ann. Statist. 8 (1980) 147–164.
[18] C.Z. Wei, Adaptive prediction by least squares predictors in stochastic regression models with application to time series, Ann. Statist. 15 (1987) 1667–1682.
[19] A. Zygmund, Trigonometric Series, 2nd Edition, Cambridge University Press, Cambridge, 1959.